AD-A016 273

GROUP DECISION ANALYSIS

N. C. Dalkey

California University

Prepared for:

Office of Naval Research
Advanced Research Projects Agency

August 1975

DISTRIBUTED BY:

NTIS
National Technical Information Service
U. S. DEPARTMENT OF COMMERCE

This research was supported by the
Advanced Research Projects Agency
of the
Department of Defense
and was monitored by the
Office of Naval Research
Under Contract No. N00014-69-A-0200-4056/452.

UCLA-ENG-7571
AUGUST 1975

GROUP DECISION ANALYSIS

N.C. DALKEY

UCLA • SCHOOL OF ENGINEERING AND APPLIED SCIENCE

47

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) GROUP DECISION ANALYSIS		5. TYPE OF REPORT & PERIOD COVERED Interim Technical Report
		6. PERFORMING ORG. REPORT NUMBER UCLA-ENG 7571
7. AUTHOR(s) N. C. DALKEY		8. CONTRACT OR GRANT NUMBER(s) N00014-69-A-0200-4056/452
9. PERFORMING ORGANIZATION NAME AND ADDRESS University of California, Los Angeles School of Engineering and Applied Science Los Angeles, California 90024		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS ARPA Order No. 2841
11. CONTROLLING OFFICE NAME AND ADDRESS Advanced Research Projects Agency 1400 Wilson Boulevard Arlington, Virginia 22209		12. REPORT DATE August 1975
		13. NUMBER OF PAGES 43
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) Office of Naval Research 800 North Quincy Street Arlington, Virginia 22217		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

This report is approved for public release. Distribution is unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Decision analysis
Group decision
Group Judgment
Delphi

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Delphi, as a procedure for aggregating judgments under uncertainty, has suffered from the lack of an underlying theoretical framework, especially one that relates group estimates to decision processes. Attempts to introduce group judment into existing theories of decision have run into difficulties exemplified by the Arrow impossibility theorem for group preferences, and an analogous theorem by the author demonstrating the non-existence of a

(over)

Block 20 (continued)

general method of aggregating probability estimates.

It is shown that consistent group preference functions can be formulated by the use of anchored scales, i.e., individual preference scales with fixed reference objects. No general resolution of the aggregation problem for probabilities appears feasible, but a justification for the use of group probability judgments can be made, based on a family of theorems to the effect that the accuracy of a group judgment is always greater than (or at worst equal to) the average accuracy of the individual judgments. Some empirical data, and some analytical results, indicate that these aggregation rules are more generally applicable, and more powerful than has been assumed in the past.

UCLA-ENG-7571
AUGUST 1975

GROUP DECISION ANALYSIS

Norman C. Dalkey

School of Engineering and Applied Science
University of California
Los Angeles, California

ABSTRACT

Delphi, as a procedure for aggregating judgments under uncertainty, has suffered from the lack of an underlying theoretical framework, especially one that relates group estimates to decision processes. Attempts to introduce group judgment into existing theories of decision have run into difficulties exemplified by the Arrow impossibility theorem for group preferences, and an analogous theorem by the author demonstrating the non-existence of a general method of aggregating probability estimates.

It is shown that consistent group preference functions can be formulated by the use of anchored scales, i.e., individual preference scales with fixed reference objects. No general resolution of the aggregation problem for probabilities appears feasible, but a justification for the use of group probability judgments can be made, based on a family of theorems to the effect that the accuracy of a group judgment is always greater than (or at worst equal to) the average accuracy of the individual judgments. Some empirical data, and some analytical results, indicate that these aggregation rules are more generally applicable, and more powerful than has been assumed in the past.

TABLE OF CONTENTS

GROUP DECISION THEORY

In the past quarter century there has been rapid progress in the theory of individual decision making under uncertainty. One of the more widely accepted points of view is that of decision analysis, or as it is sometimes called Bayesian analysis. This point of view involves the notions of subjective probability, utility, and the decision rule, maximize expected utility. (1) The theory in its present form stems from the theory of games; in fact, it can be considered the one-player version of game theory. However, it is, like the theory of games, an extension of a much older tradition concerned with rational economic decisionmaking.

In contrast, group decisionmaking has proved surprisingly intractable. Attempts to formulate a theory of group decisions have run into a spate of problems that could loosely be characterized as paradoxes of aggregation. It might be thought that a reasonable tactic would be to adopt the decision analysis framework and substitute the phrases group probability judgment, and group utility for the corresponding individual terms — in fact, this tactic has been suggested by a number of workers in the field. (2) Unfortunately, as the scatological saying has it, when this is tried, things hit the fan; troubles break out all over. Perhaps the best known of these troubles is the theorem of Kenneth Arrow which asserts that there does not exist a general method of aggregating individual preferences into a consistent group preference relation. (3) This appears to cut the foundation away from the notion of group utility. Some years ago I proved an analogous theorem showing the impossibility of a general group probability function. (4) And, as if that were not enough, even if there were no special problem with group utilities and group probabilities, difficulties can arise with the decision rule.

Figure 1 illustrates a typical difficulty of this sort. There are two individuals, i and j, who are trying to select between two courses of action, A and B. The outcome of the actions can be influenced by the events E or non-E. Each individual has his own estimate of the probabilities of the events displayed above the matrix, and each has his own assessment of the utilities of the outcome. The utilities for i are in the upper left of the boxes, the utilities for j in the lower right. The small insert boxes show the average. The value differences can be interpreted either as differences of interest — i.e., each would receive different payoffs for each outcome — or as different judgments of the value of the outcomes to the pair jointly.

Under either interpretation, both individuals think action A is preferable to action B. This is indicated by the third column, where the expected utilities for each action and each individual are listed. However, if we take the average of the two probability estimates as the group probability, and the average of the two utilities as the group utility, then the group decision would be that action B is preferable to action A.* This violates the silver rule of economic decision theory, namely the Pareto unanimity principle.** (5)

These three kinds of difficulties — with preferences, with probabilities, and with the decision rule — by no means exhaust the list of troubles that arise when group notions are introduced into decision theory. Individuals can disagree, and almost inevitably do disagree in practice, about any aspect of the decision situation. Figure 2 illustrates the simplest model of a decision

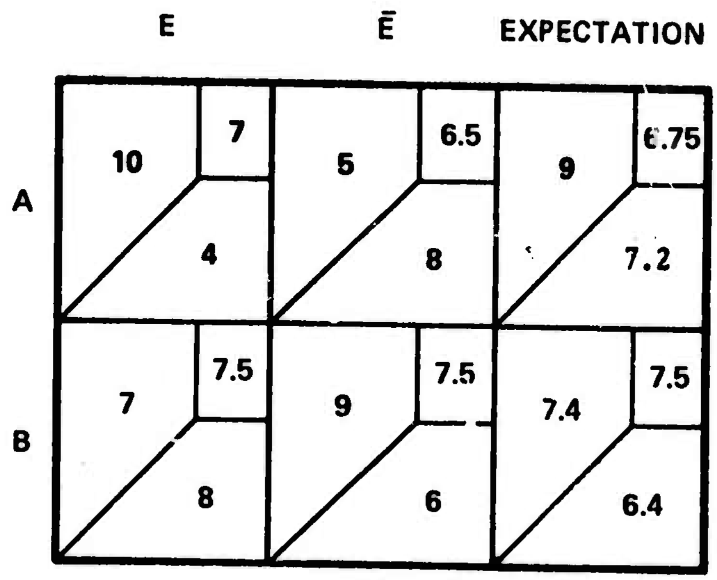
---

*The precise form of the aggregation of probabilities and utilities is not critical for the example. Other functions such as the geometric mean or the median could be used and similar "paradoxes" could be generated.

**The golden rule, of course, is <u>maximize expected utility</u>.

PARADOX OF COMPOSITION

I	.8	.2
J	.2	.8
G	.5	.5



INDIVIDUAL I PREFERS A TO B
INDIVIDUAL J PREFERS A TO B
GROUP PREFERS B TO A

Figure 1

DECISION MATRIX

	Q_1		Q_j		Q_m
	E_1		E_j		E_m
A_1					
A_i			O_{ij}		
A_n					

$V(O_{ij}) = u_{ij}$

DECISION RULE: SELECT A_i ($= A^*$) THAT MAXIMIZES $\sum_j Q_j u_{ij}$

$u_j^*(Q) = u_{ij}$ FOR A^*

Figure 2

problem; a set of potential actions A_i, a set of uncertain states of the world E_j ("E" for event) and a matrix of outcomes $|O_{ij}|$ where O_{ij} is the result of implementing action A_i when E_j is the state of the world. Individuals involved in a decision can disagree on the appropriateness of the list of actions, on the relevance of the states of the world, and on the outcomes — i.e., whether those precise consequences would indeed occur if the action were taken. The more general disagreements about the nature of the problem I have called the "point of view" issue; each individual has his own model. (6).

Most formal analyses of decisions start with the problem already formulated as a matrix as in Figure 2,* and the theory then deals with how to go on from there. Going on from there, for decision analysis means assigning probabilities to the events, assigning values or utilities to the outcomes, computing the expected outcomes of each action, and selecting the action with the highest expected value.

I will follow this procedure and assume that a statement of the decision problem in terms of a matrix is given. Each individual has his own probability distribution over the events, and his own preference relation on the outcomes. The question then becomes, from the point of view of the group, what is the best way to assign probabilities to the events, what is the best way to assign utilities to the outcomes, and what is an appropriate decision rule: Of course, the word best is just for show. We're not that far along yet.

The difficulties that arise when the decision concerns a group and the group disagrees on the relevant numbers are all of one general sort:

---

*In some versions, a more general framework, the decision tree, is used as the starting point. This more general framework is not germane to the present investigation, since all of the difficulties already show up in the simpler case of the decision matrix.

There is a set of individual judgments $\{J_i\}$, where i indexes the individual members of the group. We would like to define a function F(J), $J = (J_1,\ldots,J_n)$ which aggregates the individual judgments into a group judgment. We should like to fulfill several kinds of conditions:

1. Substantive conditions: F should be the same sort of thing as the individual judgments J_i. Thus, if the J_i are probabilities, F(J) should be a probability. If the J_i are preferences, then F(J) should be a preference relation, etc.

2. Consistency conditions: By consistency is meant coherence between the individual judgments and the group judgment. Consistency for individual judgments separately, and for the group judgment separately are presumably part of the substantive conditions. A typical consistency condition is the Pareto unanimity principle mentioned earlier; that is, if all the J_i are identical, then $F(J) = J_i$.

3. Performance conditions: If there is a figure of merit for the individual judgments, then the group judgment should do reasonably well, compared with the individual judgments, on that figure of merit. As an obvious example, if all the individual judgments are declarative sentences, and if they are all true, then F(J) should not be false.*

---

*There is some unclarity in the literature on this type of condition. For some investigators, the group process is primarily a method of arriving at a common judgment — a way of circumventing disagreement. The simple attainment of agreement is considered a sufficient good to justify "compromise" on excellence. In the common lore, a committee is expected to do rather poorly, to degrade the best capabilities of the individuals, to design a camel when you want a horse. On this view, if the group can be motivated not to design a camel, that is triumph enough. Hopefully, the following discussion will persuade the reader that more should be expected of a group than just that it not louse up the decision.

In the literature on group decision there has been little mention of conditions of type 3. There are at least two reasons: First, there is no generally accepted performance measures for preferences or values — no way to say that one individual's value judgment is correct and another's incorrect. In this respect, value judgments are ad lib. Secondly, the well-known difficulties arise from trying to meet conditions of the first two types; you can't get as far as type 3.

One thesis of this paper is that the situation can be reversed; for those cases where performance criteria exist, performance can be used to justify overlooking some inconsistencies between individual judgments. This could be called the Emerson principle.* If the aggregation procedure produces a judgment of higher excellence than the individual judgments, this fact can override some inconsistencies between the two.

A certain amount of luck enters at this stage. Since there are no performance criteria for preferences, the game would be lost if the Emerson principle were needed to get around the paradoxes of aggregation for preferences. As it happens, there is a natural resolution of the Arrow paradox without recourse to performance criteria.

Arrow's proof of the impossibility theorem is too extensive to reproduce here, but a glance at the assumptions leading to the theorem is in order.

---

* Emerson rather blew it. The quotation (from Bartlett) is "a foolish consistency is the hobgoblin of little minds...." A somewhat less restricted formulation might be: Fear of inconsistency is a hobgoblin (whether of little or big minds). A dramatic case in point is the disposition of the number zero. There was a fierce debate for two centuries on the status of zero. Accepting it as a number opened the way to contradictions, but the advantages of having zero within the pale were evident. In the end, the pragmatic side won out, with the problem of contradictions "solved" by the remarkably ad hoc rule, "don't divide by zero."

What I will contend is that there is nothing unacceptable about the intent of the assumptions; rather, it is an overstrict interpretation of the notion of ordinal which creates the problem.

The elements of the model are: (1) a set $X = \{x,y,z,\ldots\}$ of objects to be ordered.* (2) a set $I = \{1,2,\ldots,n\}$ of n individuals. (3) a set $K = \{R, R', \underline{R}\ldots\}$ of vectors of individual ordering relations over X. Each $R = (R_1,\ldots,R_n)$ consists of n individual orders. Thus xR_iy means individual i prefers x to y or is indifferent between them. A super-fixed arrow indicates strict preference, i.e., $\overrightarrow{xR_i}y$ means xR_iy and not yR_ix. (4) a function $F(R)$ which generates a group preference function over X, depending on the vector of individual preferences R.

A. Substantive conditions

1. For each R in K and each R_i in R,

   a. R_i is a complete order over X

   b. $F(R)$ is a complete order over X

2. Among the R in K there are all possible orderings by n individuals of three objects.

B. Consistency conditions

1. Monotonicity. Define $\underline{R}$ to be a forward shift of x with respect to R if: R is identical to $\underline{R}$ except for x; whenever xR_iy, then $x\underline{R}_iy$; and whenever $\overrightarrow{xR_i}y$, then $\overrightarrow{x\underline{R}_i}y$. If $\underline{R}$ is a forward shift of x with respect to R, then if $\overrightarrow{xF(R)}y$ then $\overrightarrow{xF(\underline{R})}y$.

---

*For the problem of social values, or for generating a social welfare function, X would be interpreted as states of society. However, for addressing the problem of aggregation, the precise nature of X is not germane, hence is is referred to here as "objects".

2. *Independence of irrelevant alternatives*. If R is identical to $\underline{R}$ on some subset B of X, then F(R) is identical to $F(\underline{R})$ on B.

3. *Non-imposition*. For any pair of objects x, y, there is an R in K such that $x\overrightarrow{F(R)}y$.

4. *Non-dictatorial*. For any individual i, there is a pair of objects x, y and an R such that $x\overrightarrow{R}_i y$ and $yF(R)x$.

A relation R is a complete order over a set of objects X if two conditions hold:

1. *Connexity*. For every pair of objects x, y, in X, either xRy or yRx.

2. *Transitivity*. If xRy and yRz, then xRz.

The second substantive condition requires that for at least three objects, any possible combination of individual preferences can occur, and the group preference relation is defined for all those possibilities. It is a condition to assure a certain amount of generality for the group preference function.

The first consistency condition is a sort of sure-thing principle. If x is preferred to y on the basis of a set of individual relations R, and another set $\underline{R}$ treats x at least as favorably, then surely x is preferred to y on the basis of $\underline{R}$.

The second consistency condition is a crucial one. It imposes a certain stability on the group preference. Thus if x is preferred to y by the group, and if attention is restricted to a smaller set of objects, still containing x and y, the group preference should not reverse. This is the condition that is violated by most well-known aggregation methods.

The third consistency condition is intended to assure that the group preference relation is not determined by some rule independent of the individual preferences.

The last consistency condition requires that the group preference function not be determined by the preferences of a single individual (dictator.) It asks only that for any individual, some pair of objects and some set of individual preferences exist such that the group and the individual disagree.

As I remarked earlier, the general intent of the consistency condition appears to be desirable. However, the conditions have the apparently devastating effect that there is no group preference function which fulfills them.

To see how to get out of the paradox, we need a small aside on measurement. In many discussions of measurement in economics, a broad distinction is made between ordinal and cardinal scales. The former are purely relational; if numbers are coordinated to the scale, they have only rank-order properties. In technical terms, the numbers are fixed only up to a monotonic transformation. Cardinal scales, on the other hand, have numerical properties. Several varieties of these may be distinquished (interval, ratio, etc.) depending on the degree to which the numbers are fixed by the measuring process. What is overlooked by this classification is the role of reference objects or standards. For physical interval scales such as temperature, the scale is not fixed until two different physical states have been specified — e.g., the freezing and boiling points of water at sea level — and two numbers — e.g., 0 and 100 — have been assigned to these two states. Until this coordination of numbers and physical states has been performed, the scale cannot be used to measure the temperature of a given object. For example, if an individual states that his temperature is 46, this tells you nothing until you know his reference states and his coordinated numbers for those states.

The numbers coordinated with reference objects are often called "arbitrary constants." This phaseology can be misleading. In a purely mathematical sense, the numbers are arbitrary, but that does not mean they are <u>dispensable</u>. Which states and which numbers will be employed as references can be chosen "freely" (except for practical considerations of feasibility and convenience) but <u>some</u> choice must be made before the scale becomes a measuring instrument.

Almost completely overlooked in the economic literature is the role of reference objects for ordinal scales. A typical physical ordinal scale is the Mohs hardness scale. This scale is associated with the relation <u>scratches</u>; if object x scratches object y, then x is harder than y. This is the basis of the well known test of a stone to determine if it is a "gem" by seeing if it will scratch ordinary window glass. Figure 3 shows one widely used form of the scale. Each of the ten items will scratch all of those below it. However, the associated numbers are purely ordinal — they are rank orders and nothing more. To say that the hardness of a fingernail is between 2 and 3 merely means that a fingernail will scratch gypsum and be scratched by calcite.

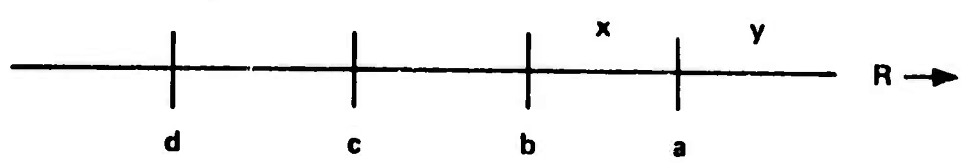
Such an ordinal scale with a fixed set of reference objects, can be called an <u>anchored</u> <u>scale</u>. An anchored scale consists of a set of objects X, a specified set of anchors A, and an ordering relation R. Usually A would be a subset of X. The scale value S(x) of an object x is the highest of the set A that has the relation R to x. As illustrated in Figure 4, $A = \{a,b,c,d\}$ and $S(x) = a$. For some purposes it may be convenient to attach numbers to the anchors, but these numbers are determined only up to a monotonic transformation.

MOHS HARDNESS SCALE

10	DIAMOND	
9	SAPPHIRE	
8	TOPAZ	
7	QUARTZ	
6	FELDSPAR	← WINDOW CLASS
5	APATITE	
4	FLOURITE	
3	CALCITE	← FINGER NAIL
2	GYPSUM	
1	TALC	

Figure 3

ORDINAL SCALE



$A_i = \{a,b,c,d\}$

$S(x) = a \qquad S(y) = m$

Figure 4

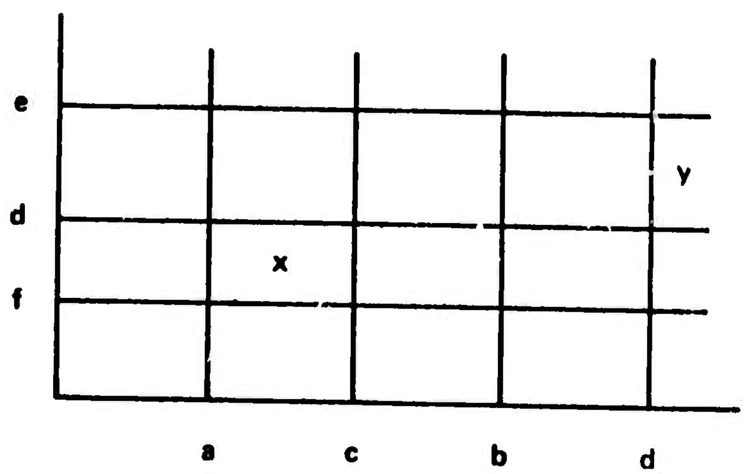
One way to interpret the Arrow theorem is: If you formulate a group preference function which disregards reference objects, it will not in general be compatible with the individual preferences. To be useful, that statement needs to be turned around to say: If individuals express their preferences in terms of anchored scales, then a group anchored scale can be formulated which fulfills the analogue of the Arrow conditions for anchored scales. This will now be investigated.

A group anchored scale can be generated from a set of individual anchored scales as follows: The anchor set for the group is the set of all n-tuples of individual anchors, i.e., the group anchor set A is the cartesian product of the individual anchor sets, $A = A_1 X A_2 X, \ldots, X A_n$. The idea is illustrated for two individuals in Figure 5. Each pair of individual anchors forms a reference point for the group. The pairs sort the objects in X into boxes, where if a and b are consecutive anchors in individual 1's scale, and c and d are consecutive anchors in individual 2's scale, the box consists of all x's such that bR_1x but not aR_1x and dR_2x but not cR_2x. The scale value of an object x is the pair of individual scale values. Illustrated in Figure 5 is the case $S(x) = (c,d)$.

There is a natural partial ordering of the objects given a group scale, namely the partial order defined by unanimity: if $S_i(x) R\, S_i(y)$ for every i, then x is preferred by the group to y. The only substantive condition not fulfilled by this partial order is connexity. What needs to be shown is that this natural partial order can be extended to a complete order without violating the analogue of the consistency conditions for anchored scales.

The group preference structure, expressed in terms of anchored scales has the elements: a set of objects X; a set of individual preference scales

GROUP ORDINAL SCALE



$S(x) = (c,d) \qquad S(y) = (m,e)$

$x \, G \, y \equiv S(x) \; G \, S(y)$

Figure 5

$K = \{S, S', \underline{S}, \ldots\}$, where each $S = (S_1, \ldots, S_n)$ is associated with anchor sets $(A_1, \ldots, A_n)$ and preference relations $(R_1, \ldots, R_n)$; a group preference function $F(S)$ associated with a group anchor set $A = A_1 X A_2 X \ldots X A_n$; and a group preference relation G. Each individual preference scale S_i is based on the associated preference relation R_i. In the group case, the order of derivation is reversed. A group preference scale is generated over the anchor set A, which then imposes a group preference relation on the entire set X. The notation designating scales and relations becomes somewhat involved. The convention will be followed that preference relations associated with scales will be represented by the quasi-arithmetic symbols $>$ and $\geq$. Differences between individual and group scales will generally be clear from the arguments. Thus $S_i(x) > S_i(y)$ states that individual i prefers the scale value of x to the scale value of y (and thus, prefers x to y). $F(S)(x) > F(S)(y)$ states that the group prefers the group scale value of x to the group scale value of y. Where no ambiguity exists, this statement will be abbreviated to $S(x) > S(y)$.

The basic modification of the Arrow conditions to make them appropriate for anchored scales are: (a) The anchor sets for all individuals are fixed, i.e., for any S_i, $\underline{S}_i$ in K, $A_i = \underline{A}_i$.* (b) The objects comprising the anchor sets are exempted from the consistency conditions. (c) For all other objects, the conditions are expressed in terms of the scale values of the objects. Thus, the modified Arrow conditions are:

---

*This does not imply that the anchor sets for different individuals are the same. In general, anchor sets for different individuals may be entirely distinct; although in practice there are obvious advantages to having common anchor sets. (a) does imply, of course, that the group anchor set is fixed.

A. Substantive conditions.

1. Each S_i in K and F(S) is an anchored scale.

2. There are three objects such that all possible orderings of their scale values by n individuals occur in members of K.

B. Consistency conditions.

1. Monotonicity. Define a forward shift of x by $\underline{S}$ with respect to S as: S is identical to $\underline{S}$ except for x. Whenever $S_i(x) \geq S_i(y)$ then $\underline{S}_i(x) \geq \underline{S}_i(y)$ and whenever $S_i(x) > S_i(y)$ then $\underline{S}_i(x) > \underline{S}_i(y)$. If $\underline{S}$ is a forward shift of x with respect to S then, whenever $S(x) > S(y)$, $\underline{S}(x) > \underline{S}(y)$.

2. Independence of irrelevant alternatives. If S is identical to $\underline{S}$ on the subset B of X, then F(S) is identical to $f(\underline{S})$ on B.

3. Non-imposed. For any x and y in X, there is an S such that $S(x) > S(y)$.

4. Non-dictatorial. For every i, there is an x, y and S such that $S_i(x) > S_i(y)$ and $S(y) \geq S(x)$.

Rather than look for conditions which guarantee the existence of a group preference scale, it is simpler to exhibit a specific group scale which satisfies the modified conditions, and thus acts as an existence proof. One appropriate scale is anchored sum of ranks. Let each individual coordinate rank-order numbers with each of his reference objects. Designate these rank-order numbers by $S_i^*(x)$.* It is convenient to let the rank order numbers start with 1 for the least preferred object. The group scale number is defined by $S^*(x) = \sum_i S_i^*(x)$. The group preference relation is defined by $S(x) \geq S(y)$ means $S^*(x) \geq S^*(y)$.

Since this procedure assigns a number to every object in X, and the arithmetic inequality is a complete order, a complete group preference order is defined on X. Monotonicity is assured since the sum is monotonic in its summands. Consistency condition 2 is fulfilled directly; the group scale value does not change when only a subset of objects is considered. Condition 3 is satisfied by invoking substantive condition 2 — there is a pair of objects x, y such that $S_i^*(x) > S_i^*(y)$ for every i — and the sum fulfills the unaminity principle. Substantive condition 2 also requires that each individual have at least two reference objects (three potential rank order numbers) and hence non-dictatorship is fulfilled. There is a pair of objects x and y such that $S_i^*(x) = S_i^*(y) + 1$, but $S_j^*(y) = S_j^*(x) + 2$ for $j \neq i$. Hence $\Sigma S^*(y) = \Sigma S^*(x) + 2(n-1) - 1$. Thus x is preferred to y by individual i, and y is preferred to x by the group.

This completes the demonstration that anchored sum of ranks fulfills the analogues of the Arrow conditions for group preference scales, and is thus an existence proof for group preference functions.**

---

*This will not work if the anchor set is infinite at both ends, or if different individuals have anchor sets infinite in different directions. There is no problem dealing with infinite anchor sets, but they are overlooked here because the essential difficulties expressed by the Arrow theorem arise with finite sets.

**There may be some uneasiness that anchored sum of ranks is not purely ordinal in the sense that the group function depends on the numerical values of the rank order numbers. Thus, if one individual multiplied all his rank order numbers by some large constant, he would become an arithmetic dictator. This objection misconstiues the role of the rank order numbers for the existence proof. They are simply a device to define a group scale which is consistent. Notice that once this group scale has been defined, the rank-order numbers can be "thrown away" and the group scale applied in a purely non-numerical fashion.

Anchored sum of ranks is just one out of an infinite number of consistent group scales that can be defined. In a way this is disappointing. The selection of a specific group function in practice would depend on other properties than those contained in the Arrow conditions.

Aside on Electing a President

As is well known, the type of difficulty expressed in the Arrow theorem has serious implications for all group decisions involving voting-like procedures. The most serious are the dominating role of the agenda when sequential (progressive elimination) techniques are used (7) and the "spoiling" effects of "irrelevant" candidates. In the French style of election where there is a runoff between the two leading contenders if there is no majority candidate, there are many plausible "scenarios" which suggest that the candidate most highly rated by the total electorate can be eliminated on the first round. It is even easy to design situations in which the least preferred candidate out of three is elected (c.f., the U.S. example below.)

In the United States, the situation is obscured by the electoral college, and the fact that there are usually only two major candidates. However, the issues still lurk in the background. Consider, for example, the election of 1912, with Wilson, Taft, and Roosevelt as the three major candidates. We don't have a record of voter preferences among these, just the record of first preferences. A plausible assumption would be that most of those who voted for Taft or Roosevelt would have preferred either to Wilson, and those who voted for Wilson would have preferred Roosevelt to Taft. There assumptions generate the preference table which follows.

	Wilson	Roosevelt	Taft	Number $(\times 10^6)$
Wilson	1	2	3	6.3
Roosevelt	3	1	2	3.5
Taft	3	2	1	4.2

Straight majority vote on this table would lead to the preference order Roosevelt-Taft-Wilson. Sum of ranks (weighted by numbers of voters) gives the order Roosevelt-Wilson-Taft. In either case, Roosevelt is the "preferred" candidate, and in the case of majority vote, Wilson is the least preferred.

This type of mis-selection could be eliminated if anchored scales were used. In the case of the U.S. presidential elections there is a natural set of anchors, namely, the list of all past presidents. A plausible voting scheme would be to have each voter rank-order all the past presidents in terms of his perception of their desirability as presidents. This could be done at the voter's leisure at any time between elections. There is no necessity that the rank orders of any individual agree with those of any other.* At election time, each voter casts his ballot by reporting the position in his scale of each candidate. The candidate receiving the highest sum of ranks is elected.

The scheme will work for as many candidates as the voters have time to rate. It has the side benefit that the final tally would give a fairly diagnostic reading on the voters evaluation of the candidates.

There is a possible weakness in the procedure as described. A significant segment of the voting public might attempt to bias the ratings by, for example, giving the highest possible rating to their favorite candidate, and

---

*Since there are 38 presidents, there are $38! = 5.23 \times 10^{44}$ permutations, which is quite enough for each voter to have a different ordering!

rating all the others at the lowest level. This would vitiate the procedure. There is a simple way around this difficulty, one that is perhaps a little cumbersome, but not without attractions of its own. The resolution is affected by starting with a large slate of initial candidates — say 50 for purposes of illustration — all of which are rated by the voters. After all ratings are in, a small final slate — say 5 — are selected at random. The candidate in this final slate with the highest sum of ranks would then be declared president. The numbers 50 and 5 are just illustrative. Some statistical engineering could be done to determine the minimal sizes for the two slates keeping to an acceptable level the probability that the finalists were not all from the bottom of the heap. I would imagine that a lottery of the type suggested would be a dramatic event. It should have a very high rating if telecast live.

The question whether the procedure would be feasible for the "average citizen" doesn't appear very serious. It would require somewhat more background and a little more time than now appears to be devoted to voting by the electorate.

The rank order scale is itself relatively crude, and could probably be improved upon. However, this is a second order consideration (especially with 70 or so million voters) compared to the stability and consistency afforded by the anchored rating procedure.

Note on Numerical Utilities

Once having found that consistent group preference functions can be generated, there is no obvious reason why the advantages of cardinal utility functions should be exploited. The subject is treated much more fully elsewhere (8). I will content myself with two points.

If it is assumed that each individual member of the group has a numerical utility function on the set of objects X (e.g., of the sort elaborated by von Neumann and Morgenstern, where the scale is determined up to a linear transformation (9)) then individual reference sets need contain only two objects. This is a great simplification over ordinal scales where a large set of reference objects might be needed to determine the individual scales with sufficient precision. More significant is the fact that reference sets for general social value scales are difficult even to imagine — most individuals have not had enough experience with enough states of society to designate a well-defined set of "objects." The assumption that each individual rates social states solely in terms of his own consumption appears to be a radical oversimplification. Although the assumption that each individual has an interval utility scale on states of society also appears to be highly unrealistic, some of the implied conditions for social utility scales might be more palatable than the assumption that society could examine the individual anchor sets of large numbers of individuals and select a social ordering of the cartesian product.

Under the assumption of individual cardinal utility scales if any of several elementary additional assumptions are made, the form of the group utility function becomes sharply restricted. For example, if the assumption is made that when the group finds two objects x and y equivalent, then it is indifferent between either and any probability mixture of the two, then the group utility function takes the form of a weighted sum of the individual utilities; i.e., if U_i is the utility function for individual i, and U is the group utility function, then $U = \Sigma w_i U_i$. The w_i in this case perform a dual role of rescaling each individual utility to conform to the others, and also of determining the proportionate share of each individual in social benefits.

Although a number of objections have been raised against the linear social utility function, it has some strong advantages. This is especially true if it is assumed that the opportunity space (space of achievable outcomes) is concave, in which case some of the more salient criticisms become "academic" — i.e., are concerned with cases which are not likely to arise.

If it is assumed that an absolute zero can be defined for individual utilities — possibly complete destitution guaranteeing death — then a multiplicative form for the group utility looks attractive. In symbols, $U = \prod_i U_i^{w_i}$. Here the weighting factors appear as exponents. As John Nash pointed out long ago, the product has the desirable feature that it is invariant under multiplicative transformations (10), and hence, given the assumption of an absolute zero, invariant under all permissable transformations. Unfortunately, the product is not compatible with the assumption of unamimity on probability mixtures.

Performance Criteria for Probabilities

The situation with group probability estimates is quite different from that with group preference judgments. It looks very unlikely that any "natural" resolution of the inconsistencies between individual and group estimates can be found.* The reason is that the constraints on probabilities are much more severe than those on preferences. In particular, probabilities are fixed numbers allowing no tranformations; i.e., if p is a probability measure on a set of events, there is no function $f(p) \neq p$ which is also a probability measure on the same set of events. For group estimates, the only

*I say unlikely, rather than impossible because there is the outside chance that some measure of uncertainty other than probability will turn out to be both a reasonable way to express incomplete information, and will aggregate in a consistent fashion.

identity function is the dictatorial one, $f(p_1, p_2, \ldots, p_n) = p_i$, where i is a given individual.

There are no dramatic paradoxes which arise from this situation. Simple illustrations of the type of difficulty: The average of a set of probabilities fulfills the requirement that probabilities of exclusive events add; however, it does not fulfill the requirement that the probability of the conjunction of two independent events is the product. The converse is true for the product as an aggregation rule — it does not sum to one for exclusive and exhaustive events but is multiplicative for conjunctions.

If there is any hope of "rescuing" group probability estimates from inconsistency, we apparently need to invoke the Emerson principle. This requires specifying a figure of merit for probability estimates. In the past decade or so there has been a rapid development of a theory of probability assessment which furnishes an appropriate criterion.

There are several directions from which this theory can be approached. One of the most perspicuous, if not perhaps the most profound, begins with the desideratum of keeping the estimator "honest." The theory consists of a reward scheme which will motivate the estimator to report what he believes to be the relevant probabilities. Several basic notions are needed to expound the idea.

$\{E_j\}$ A set of (exhaustive and exclusive) events for which probabilities are desired.

$\{Q_j\}$ The probabilities on E_j which the estimator believes.

$\{R_j\}$ The probabilities which the estimator reports.

$\{P_j\}$ The (unknown) objective probabilities*

$S(R,j)$ A reward function which, after the fact, pays the estimator an amount S, depending on the report R, and the event j which occurs.

To say that S rewards the estimator for being honest is to say

$$\sum_j Q_j S(R,j) \leq \sum_j Q_j S(Q,j)$$

That is, the estimators' (subjective) expected reward is greatest when he (honestly) reports what he believes. There is a large class of functions which fulfill this condition. These have been extensively studied (12, 13, 14). Among the better known are the logarithmic scoring rule, $S(R,j) = \log R_j$ and the quadratic scoring rule, $S(R,j) = 2R_j - \sum_j R_j^2$. It is easy to see that the sum of any two scoring rules is a scoring rule, and any linear transformation, aS + b, where a and b are constants, is a scoring scheme. Various names have been given to these reward structures — reproducing score, admissible score, probabilistic score, proper score, honesty score, etc. I will use the shortest — proper score.

There are a number of properties of proper scores which can be derived fairly directly from the definition. S rewards the estimator not only for being honest, but also for being accurate; i.e.,

$$\sum_j P_j S(R,j) \leq \sum_j P_j S(P,j)$$

This follows immediately from the definition by substituting P for Q. Thus, the objective expected score is a maximum when the estimator reports the objective probability.

---

*There is some dispute whether objective probabilities can be defined for all types of estimates of interest in decision theory. Rather than arguing the point here, I simply examine the consequences of assuming that there is an objective probability. For a fuller discussion see (11).

A proper score rewards the estimator for being precise, i.e., for reporting probabilities close to 0 or 1. This results from the fact that $\sum_j Q_j S(Q,J)$ is convex. (15)

A proper score can be thought of as an extension of the notion of truth-value to the case of probabilistic estimates. For declarative assertions — "It will rain tomorrow" — the score is two-valued, true (or 1) if the event occurs, false (or 0) if the event does not occur. For probabilistic statements — "The probability of rain tomorrow is p" — the score is S(p, rain) if it rains and S(p, not-rain) if it doesn't rain. The two-valued scheme has an analogue among proper scores, namely, the score rule that pays 1 if the event with maximum reported probability occurs, and 0 otherwise. In a sense, this is the score rule used in grading objective examinations, if we assume that the student checks the alternative that he thinks has the highest probability of being true.

It is convenient to divide proper scores into two sorts: informational and economic. Informational scores are those which depend only on the reported probabilities and the event that occurs and on no other properties of the situation. Economic scores depend not only on the reported probabilities but also on the decision situation, e.g., on the payoff resulting from a decision.

Among the informational scores, there is a special group which have been considered the most appropriate for scientific studies, and might be labeled scientific scores. These have a property that can be called exactness, i.e., the scores motivate the estimator to furnish exact reports of his beliefs. The two-valued score mentioned above motivates the estimator only to report a higher probability for the event he thinks most likely than for the others.

An exact score clearly must have a continuum of values. The logarithmic and quadratic scores mentioned above are exact. Most of the scientific scores have an important additional property; namely, $S(R,j)$ is concave in R.

Informational N-heads Rules

One way to express the Emerson principle for probability estimates is to say that the group will perform better, in terms of prababilistic scores, than the individual members of the group. Given a set of estimates $\{Q_{kj}\}$ by a group (k indexes individuals), the average objective expected score is

$$OES = 1/n \sum_k \sum_j P_j S(Q_{k,j}) = \sum_j P_j 1/n \sum_k S(Q_{k,j})$$

I have assumed each individual is honest and reports his believed probabilities Q_k. In the more interesting cases, P is unknown, and the average objective expectation cannot be computed. However, we can ask, under what circumstances is the average expected score of the individuals less than the expected score of the group; i.e., when is OES less than $\sum_j P_j(\bar{Q},j)$ where $\bar{Q} = 1/n \sum_k Q_k$, independently of P and $\{Q_k\}$? It is not difficult to show that a necessary and sufficient condition for the inequality to hold for all P and $\{Q_k\}$ is that $S(Q,j)$ be concave in Q.

Hence, for those scientific probabilistic scores which are concave, such as the log score and the quadratic score, the result holds that the objective expected score of the group will always be greater than or at worst equal to the average expected score of the individuals. Over a large number of estimates, the observed total score of the group should be larger than the average total score of the individual members.

I call a statement to the effect that a group judgment receives a higher performance rating than the average rating of the individual judgments an n-heads rule (generalization of the adage "two-heads are better than one.") The elementary n-heads rule enunciated above is just one of a large family of such rules, where the precise form of the rule depends on the kind of estimate, on the scoring rule, on the aggregation rule for individual estimates, and on the kind of expectation employed (absolute,* objective, or subjective.)

Somewhat more definitive n-heads rules can be derived if the method of aggregation is tailored to the form of score rule. For example, the geometric mean "fits" the logarithmic score rule better than the mean. Thus, it is shown in (16) that the objective expected log score of the geometric mean is precisely equal to the average expected score of the individuals plus a term D which is a function of the dispersion of the individual estimates but is independent of the objective probabilities. The higher the dispersion, the greater D — i.e., the greater the advantage of the group score over the average individual score.

The various n-heads rules would appear to furnish a justification for the utilization of group probability estimates, even if there is some inconsistency between the group estimate and the individual estimates.

Economic N-heads Rules

The results of the previous section concern a small subclass of proper scoring rules, namely those that are concave. For many decisions, the most appropriate performance criterion is the payoff as defined in the decision matrix. This measure does not in general lead to concave functions.

Define an enterprise as a group of individuals who are faced with a decision matrix as in Figure 2. Various sorts of enterprises can be

---

*Absolute means non-probabilistic, a type of rule not examined in this paper.

distinguished, depending on how the group wishes to proceed, and the degree of commonality assumed for utility functions. The simplest type of enterprise is one where the individual utility functions coincide, and the group has predetermined that they will select one common action. This type of enterprise could arise from the group having established a group utility function with the rule that all members will attempt to maximize this function. An analogous case arises in the more familiar situation of an economic partnership, where the group utility is just the proceeds of the firm, and each member receives a proportionate share of the proceeds.

We first establish a general result, namely, that any decision matrix, with a given utility function, and the decision rule maximize expected utility, is a proper scoring rule for estimates of the probabilities. Let $\{Q_j\}$ be an estimate of the probabilities for a decision matrix $|U_{ij}|$. The expected utility of action A_i as a function of Q, $U_i(Q)$, is $\sum_j Q_j U_{ij}$. We define $U^*(Q,j)$ as U_{ij} of the action A_i for which $U_i(Q)$ is a maximum. Thus $\sum_j Q_j U^*(Q,j)$ is the maximum achievable expected utility, given Q. It follows from the definition that

$$\sum_j QjU^*(Q,j) \geq \sum_j Q_j U^*(R,j)$$

This inequality has precisely the defining form for a proper score rule, where $U^*(Q,j)$ plays the role of $S(Q,j)$.

This score rule has sometimes been called the "piece of the action" rule — to be applied to a consultant, for example, who is advising a firm by furnishing estimates of probabilities for relevant contingencies. (17) We are applying it more generally to the case of all concerned individuals, whether consultants or members of the firm, where the payoff is some proportion of the proceeds of the firm. Raiffa has called the rule in this context the "naturally imputed score rule." (18)

In the simplest case there is an agreed-on rule that a single action will be taken. There is no loss of generality in assuming that this action is one which is optimal for a given estimate R of the probabilities.* The average expected payoff to the enterprise as perceived by the members of the group will be

$$EU = 1/n \sum_k \sum_j Q_{kj} U^*(R,j) = \sum_j 1/n \sum_k Q_{kj} U^*(R,j) = \sum_j \bar{Q}_j U^*(R,j)$$

where $\bar{Q}_j = 1/n \sum_k Q_{kj}$. Since $U^*(R,j)$ is a proper score rule, $EU \leq \sum_j \bar{Q}_j U^*(\bar{Q},j)$.

This is the simplest n-heads rule for an economic scoring scheme. It can also be taken as a formulation of an informational n-heads rule, where the reward function is not concave. Here the relevant criterion is not the objective expectation, but the average subjective expection -- the expectation based on the beliefs of the members of the group. This result, although not as strong as obtained with concave score rules, nevertheless is still fairly impressive. It states that, even for an enterprise where the payoff may be specified in terms of "cold cash," if the members of the enterprise disagree on the relevant probabilities, then the expected payoff of that enterprise, based on a group estimate of the probabilities, will be higher than the average expected payoff predicted by the individuals.

This may not satisfy every member of the group, since it is clear that each individual thinks the enterprise would do better if it followed his advice. We can explore this a little further. Suppose we introduce the notion of the <u>Monday-morning-quarterbacking-payoff</u> (MMQP) as follows: Irrespective of what the enterprise does, each individual is paid, after the

---

*This rules out the trivial case where an action might be chosen which is dominated by some mixture of other actions.

fact, some fraction of what the enterprise would have made if it had followed his advice. Without going into niceties here, since we are dealing with expectations, we will let the phrase "what the enterprise would have made" be defined by the decision matrix. Thus, each individual k is paid $U^*(Q_k,j)$, where U^* is defined by the optimal action given Q_k and j is the event that happens.

Individual k sees the total group as receiving

$$\sum_{\ell}\sum_{j} Q_{kj}U^*(Q_{\ell},j)$$

Taking the average of these perceptions, we have

$$1/n \sum_{k}\sum_{\ell}\sum_{j} Q_{kj}U^*(Q_{\ell},j) = \sum_{\ell}\sum_{j} \bar{Q}_j U^*(Q_{\ell},j) \leq n \sum_{j} \bar{Q}_j U^*(\bar{Q},j)$$

since U^* is a proper score rule.

Even in this disaggregated case, where we have "every man for himself" to begin with, the average expectation of total group return is maximized by each individual adopting the same (average) group estimate. This formulation can be made more realistic by assuming the group agrees beforehand to pool their earnings and redivide after being paid. An elementary example might be a group who agrees to engage in a series of gambling ventures. Each makes his own bets, but the proceeds are pooled. Their average expectation will be maximized if they decided beforehand to use a group predicion concerning the outcome of each gamble.

The economic n-heads rule can be extended to the case of a non-common payoff, retaining the assumption that a common action will be taken. However, the story is a little monotonous — almost any way you view an enterprise, if there is disagreement on probabilities or utilities, but agreement on the rule of common action, the expectation of the group judgment is greater than the average expectation of the individuals.

Empirical Validation

Most of the results presented so far in this paper are methematical and have limited empirical content. Given that individual utilities and probability estimates fulfill the standard substantive conditions, the n-heads rules follow tautologically.

Nevertheless, there is an understandable reluctance to put complete trust in such formulations for real life decisions. The desire to see them "tried in practice" is strong, and I think justified, even though it is difficult to specify exactly what the issue is. The Missouri rule "show me" has a good, final ring to it. In part, this impulsion comes from the overall simplifications and extrapolations that are a natural part of mathematical models. Although each simplification may seem justifiable separately, there is a reasonable sense in which it can be asked whether every-day decisions are expressed sufficiently well by the standard decision matrix so that the predictions of theory can be trusted.

Unfortunately some of the most interesting results, especially those concerning economic n-heads rules, were generated only within the last few months, and there has not been sufficient time to carry out relevant experiments. Most of the experimental studies relating to group judgment have been conducted within a different conceptual framework. However, it is worth trying to see if some previous experimental results can be interpreted in light of the present analysis to give an initial empirical back-up to the theory.

A first look suggests a rather surprising possibility. The results of at least two studies concerning betting appear to support an even stronger n-heads rule than any derived in the previous sections. This result is that the <u>observed</u> payoff for the group estimate is higher than the <u>observed</u> average payoff over individuals. Although the theory does not reject this result for any given experiment, it does not predict it. The result cannot be derived from the elementary fact that a decision matrix is a proper score rule. In the case of a bet, we have the decision matrix illustrated in Figure 6.

	E	not-E
A. Bet on E	$\frac{1-u}{u}$	- 1
B. Bet on not-E	-1	$\frac{u}{1-u}$

Figure 6

Payoff matrix for simple bet,
(standard bet of 1 unit)

where 1-u/u are the appropriate odds for a positive bet on an event with probability u. Maximization of expected payoff would require selecting A if the individual's belief was that the probability of E is greater than u, otherwise B. The derived score rule for this matrix is not concave, and in general, the average objective expected score for a group is not necessarily less than the objective expected score of the group average — it depends on the unknown objective probabilities. For example, for a group of two, with $u = .4$, if the objective probability is $p = .6$ and individual one thought the probability of E was .5 and individual two thought the probability was .2, then the average of the probabilities is .35, which would lead to a bet on B. The group expected payoff would be -.33, whereas the average expected payoff would be .083.

The published study by Robert Winkler is an experiment with bets on football games by graduate students and faculty at the University of Indiana.[19] The study was concerned primarily with assessing the probability estimates of the subjects in terms of informational score rules, but includes the performance in terms of monetary payoffs for hypothetical bets. Though hypothetical, the bets were realistic in the sense that if they had been placed the computed payoffs would have been realized.

The relevant results of this study are presented in Table I. The outcomes are expressed in terms of net gain per dollar bet.

Table I

	Bets on Big Ten Games	Bets on NFL Games
All subjects	-.119	-.091
Consensus	-.094	-.031

Winkler adds, "Moreover,... a consensus consisting of the faculty subjects alone ... did even better."

If a different betting strategy was employed, namely one where the amount of the bet depended on the point spread quoted by the bookie, in this case Bet = $(E-B)^2$ where E is the individual's expected point spread computed from his probabilities, and B is the bookie's reported point spread, the results are even more dramatic.

Table II

	Big Ten	NFL
All subjects	-.179	-.085
Consensus	.291	-.011

These results are similar to an unpublished study conducted at the RAND Corporation in the early exploratory phase of the group judgment project. In this case, the group was a group of horse-race handicappers, and the comparison was between bets placed on advice of individual handicappers and those based on the majority vote of the handicappers. The results were similar to those in Table I, the group advice lost less money than the average individual advice. At that time this was taken to be a negative result, hence the study was not published!

It is difficult to compose a meaningful null hypothesis for these two studies; thus it is hard to assess the significance of the better performance of the group over the average performance of the individuals. Winkler's study appears to be large enough to rule out "simple chance."

One possibility suggested by these results is that there is a basic difference between a <u>single bet</u> and <u>repeated bets</u> with a wide distribution of odds. This observation receives some support from the gambling-house model employed by Brown as a device for generating scoring rules.[20] Although Brown uses the model as a "gedanke experiment," it can be reformulated to have a more literal interpretation. Suppose a group of individuals experience a succession of betting opportunities, each expressible by the matrix

	E	not-E
A. Bet on E	1/u	0
B. Bet on not-E	0	1/1-u

Figure 8

Strategically Equivalent Matrix for Simple Bet

This is obtained from Figure 7 by adding 1 to all entries, giving a strategically equivalent matrix.

The sequence of opportunities can be characterized by a distribution D(u) of the parameter u, $0 \leq u \leq 1$, which determines the odds offered. To complete the model, we must assume independence between the believed probabilities Q_k of the members of the group and the parameter u. The decision rule, select A if $Q_k > u$, otherwise B, leads to a variety of expected payoffs, depending on the distribution D(u).

$$\text{Expectation if E occurs} = \int_0^p \frac{D(u)}{u}\, du$$

$$\text{Expectation if not-E occurs} = \int_p^1 \frac{D(u)}{1-u}\, du$$

It is easy to see that the expected payoff is a proper score rule, since the decision rule is a proper score rule for any given u, and the sum of a set of score rules is a score rule.

For some distributions D(u), the expected payoff is, in fact, concave in Q. For example, if D(u) is uniform between 0 and 1, the expectation is the logarithm. If $D(u) = ku(1-u)$, the quadratic rule results. The latter distribution is rather appealing, since it assumes that opportunities with extreme odds (u close to 0 or 1) are relatively rare. However, higher order distributions of the form $ku^r(1-u)^s$ do not generate concave expectations. (21)

Tabulating available odds for various kinds of gambling situations would quickly show which have distributions that are favorable for objective n-heads rules. There is clearly a rich area of investigation possible here, both empirical study of distributions of opportunities, and analytic study of appropriate distributions for various sorts of decision matrices.

Coda

The foregoing does not add up to a complete theory of group decision. Rather it presents a framework within which certain perceived difficulties with group decision can be resolved. Thus, inconsistencies between individual and group preferences can be dealt with by anchored scales. Inconsistencies between individual and group probability estimates can be adjudicated by showing that group estimates will furnish higher performance scores than the average of individual scores.

In any given decision situation, selection of a specific group utility measure or a specific probability aggregation technique requires considerations not contained in the framework. Of course, there are some hints. For many purposes, simple additive functions would appear to be acceptable approximations.

For those social processes where group decisions are now in use (or are desired), the group decision analysis framework offers a wider and more coherent set of procedures than now commonly used. In addition, the economic n-heads results suggest that group decisions have a broader scope and greater power than has been assumed. It seems likely that group procedures would demonstrate advantages in many contexts which at present are the province of individual decisionmakers.

NOTES AND REFERENCES

1. Vide, Howard Raiffa, Decision Analysis, Addison-Wesley, Reading, Mass., 1968, or Ronald Howard, "The Foundations of Decision Analysis," IEEE Transactions on Systems Science and Cybernetics, Vol. SSC-4, No. 3, September 1968, pp. 211-219.

2. This suggestion has been made by a number of contributors to decision analysis, including Harold Raiffa, and Ward Edwards. Cf. Ralph L. Keeney and Craig W. Kirkwood, "Group Decision Making Using Cardinal Social Welfare Functions," Technical Report No. 83, Operations Research Center, MIT, October 1973.

3. Arrow, Kenneth, Social Choice and Individual Values, John Wiley and Sons, New York, 1951

4. Dalkey, N., "An Impossibility Theorem for Group Probability Functions," The Rand Corporation, P-4862, June 1972.

5. This example is similar to one discussed in Raiffa, op. cit.

6. This topic is explored in the Introduction to Dalkey, N., et al., Studies in the Quality of Life, D. C. Heath, Lexington, Mass, 1972.

7. Plott, Charles R., and Michael E. Levine, "On Using the Agenda to Influence Group Decisions: Theory, Experiments and Application," Presented at the Interdisciplinary Colloquium, Western Management Science Institute, UCLA, January 1974.

8. Dalkey, N., Group Decision Analysis, forthcoming, c.f., Keeney, op cit., and L. S. Shapley and M. Shubik, "Game Theory in Economics — Chapter 4: Preferences and Utility," The Rand Corporation, R-904/4-NSF, December 1974.

9. von Neumann, J., and O. Morgenstern, Theory of Games and Economic Behavior, Princeton University Press, Princeton, N.J., 1947.

10. Nash, J.F., "The Bargaining Problem," Econometrica, Vol. 18, No. 2, April 1950.

11. Dalkey, N., "Toward a Theory of Group Estimation," in Linstone, H., and M. Turoff, The Delphi Method: Techniques and Applications, Addison-Wesley, Reading, Mass., 1975.

12. Savage, L. J., "Elicitation of Personal Probabilities and Expectations," J. Amer. Stat. Assoc., Vol. 66, December 1971, pp. 783-801.

13. Brown, Thomas, "Probabilistic Forecasts and Reproducing Scoring Systems," The Rand Corporation, RM-6299-ARPA, July 1970.

14. Shuford, Emir H., Jr., Albert Arthur, and H. Edward Massengill, "Admissible Probability Measurement Procedures," Psychometrika, 31, June 1966, pp. 125-145.

 Preceding page blank

15. Cf., Savage, op. cit.

16. Dalkey, in Linstone and Turoff, op. cit.

17. Savage, op. cit.

18. Raiffa, H., "Assessments of Probabilities," unpublished, 1969.

19. Winkler, R., "Probabilistic Prediction: Some Experimental Results," J. Amer. Stat. Assoc., Vol. 66, December 1971, pp. 675-685.

20. Brown, T., op. cit.

21. Judea Pearl has obtained cognate results working with a different decision matrix, namely

	E	Not E
A	0	x
B	y	0

and a different formulation of the distribution function, namely identical, independent distributions on x and y, so that $g(x,y) = f(x)f(y)$. In some respects this is a more general formulation than the gambling model since x and y are arbitrary payoffs. On the other hand, the assumption of independence is a definite restriction. The form of the distribution f that generates a given scoring scheme is, of course, quite different from the distribution $D(u)$ that generates the same scoring scheme for the gambling model. Vide, Pearl, Judea, "An Economic Basis for Certain Methods of Evaluating Probabilistic Forecasts," UCLA-ENG-7561, School of Engineering and Applied Science, University of California at Los Angeles, July 1975.